Supplementary Material

Association of a lincRNA *postmortem* with suicide by violent means and *in vivo* with aggressive phenotypes

Giovanna Punzi M.D., Ph.D., Gianluca Ursini M.D., Ph.D., Giovanna Viscanti Psyc., Eugenia Radulescu M.D., Ph.D., Joo Heon Shin Ph.D., Tiziana Quarto Ph.D., Roberto Catanesi M.D., Giuseppe Blasi M.D., Ph.D., Andrew E. Jaffe Ph.D., Amy Deep-Soboslay M.Ed., Thomas M. Hyde M.D., Ph.D., Joel E. Kleinman M.D., Ph.D., Alessandro Bertolino M.D., Ph.D. and Daniel R. Weinberger M.D.

Supplementary Table 1

	Non Suicide	*Suicide by Non-Violent Means*	*Suicide by Violent Means*	*Total*	*Statistics*[2]
N	101	50	77	228	
Age[1]	48.45±13.9	44.25±13.52	41.62±14.47	45.22±14.28	Suicides are younger than non suicides (p=0.002), especially suicides by violent means (p<0.002)
Males	61	25	58	144	Among suicides, more men choose violent means than females (p=0.006)

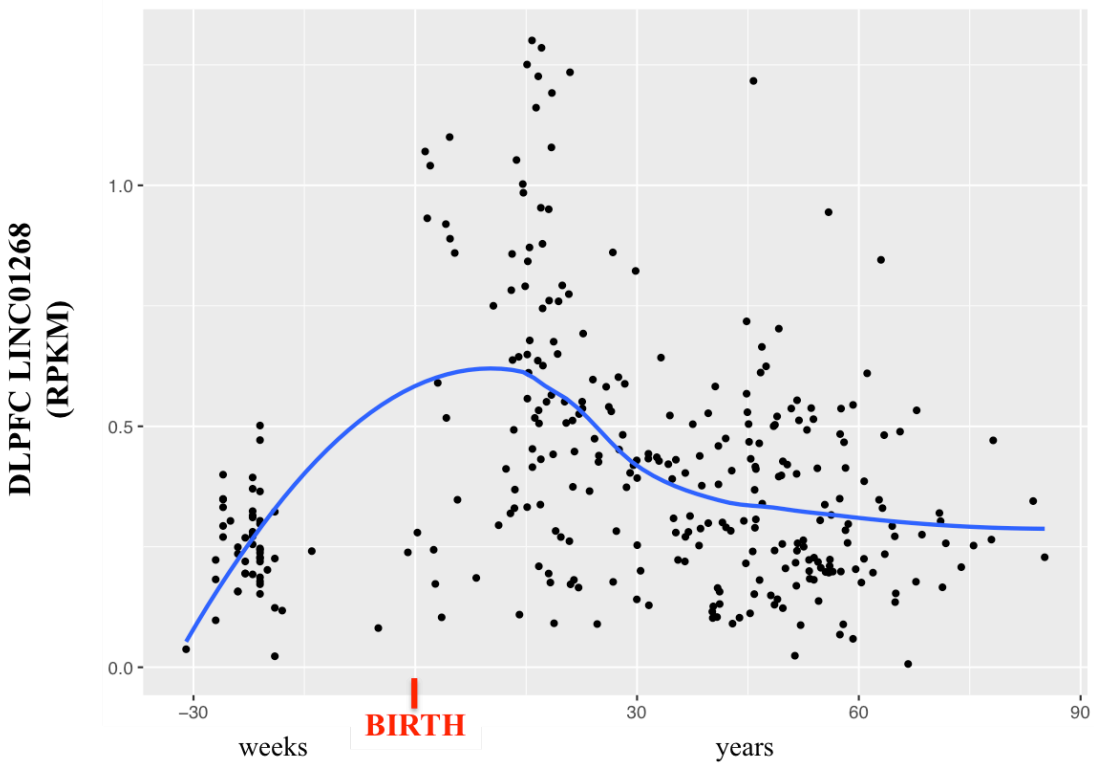
[1]Values are years, means ±SD
[2]Two sample t-test and X_2 test for difference

	SCZ	*MDD*	*BPD*
Non Suicide	51	36	14
Suicide by Non-Violent Means	11	24	15
Suicide by Violent Means	16	47	14

ST1: The combined patients sample, containing the previously reported subjects and the new sample. The tables present (top) demographics for the patients sample and (bottom) diagnosis. SCZ=schizophrenia; BPD=bipolar disorder; MDD=major depression. All brains have RIN ≥6.9; mean pH from cerebellar tissue is 6.42 (SD±0.26), and mean postmortem interval (PMI) is 32.73 hours (SD±20.63).

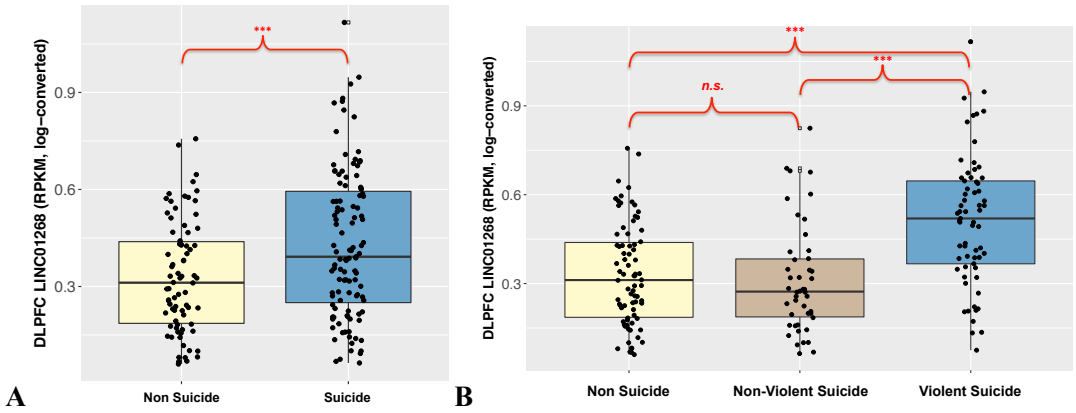
Supplementary Figures

Supplementary Figure 1



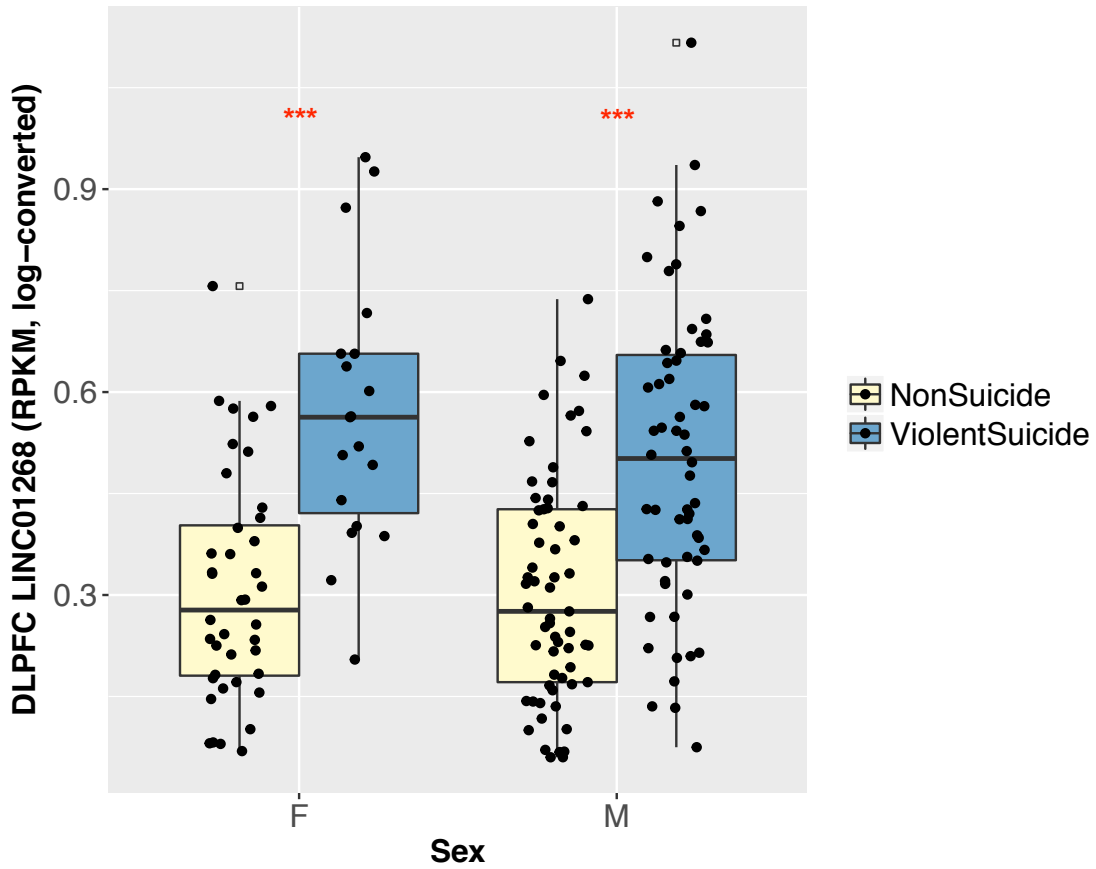
SF1: *LINC01268* DLPFC expression in a sample of normal controls of varying ages and races. Peak expression of the gene appears to be in early adult life.

Supplementary Figure 2



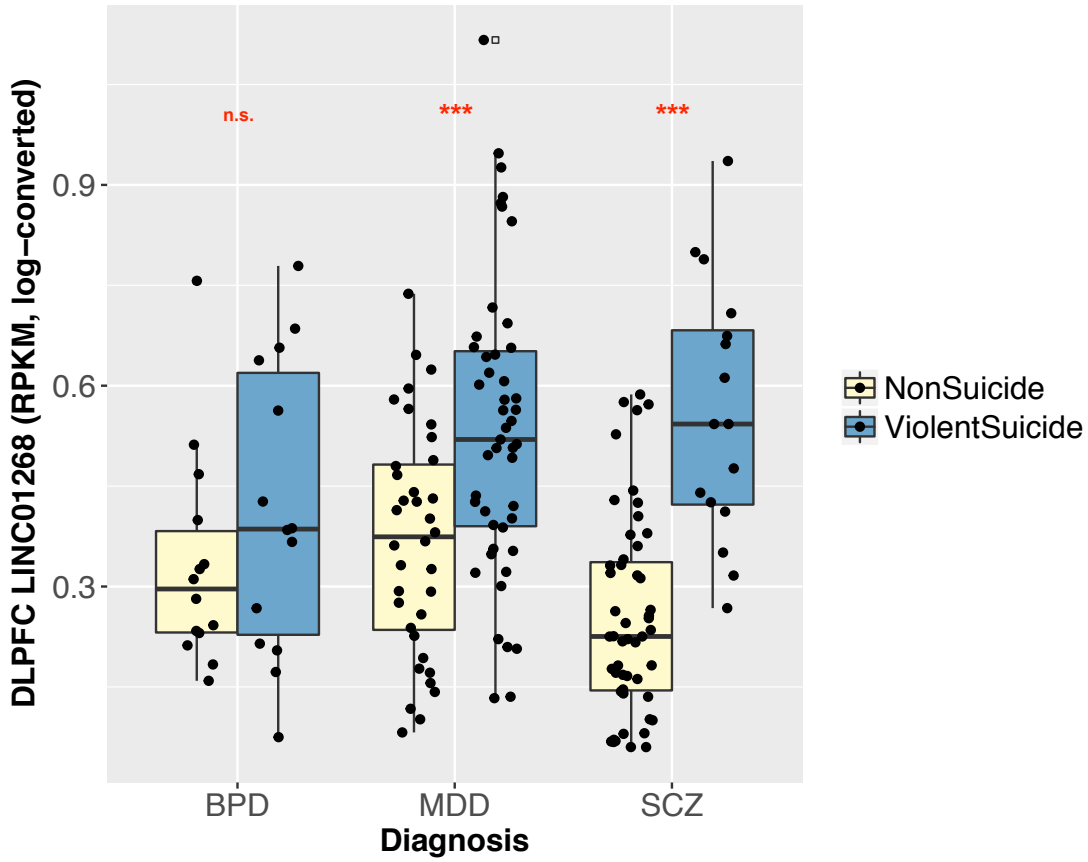
SF2: Boxplots of the effect of manner of death on the DLPFC expression of *LINC01268* in the replication sample (N=189, All PZ, with age, sex and RIN as covariates). Manner of death is significantly associated with *LINC01268* expression, which is greater in suicide completers compared with non-suicidal deceased **(A)**, and specifically in suicides by violent means (N=65) compared with non-suicides (N=78) and suicides by non-violent means (N=46) **(B)**. Non-suicides and suicides by non-violent means do not differ. Statistics: **(A)** non suicide vs suicide: t=2.750, p=0.0065 (df 1, 184); **(B)** non suicide vs violent suicide: t=5.064, p=1.29e-06 (df 1, 138); non suicide vs non-violent suicide: t=-0.505, p=0.61433 (df 1, 119); non-violent suicide vs violent suicide: t=5.116, p=1.4e-06 (df 1, 106).

Supplementary Figure 3



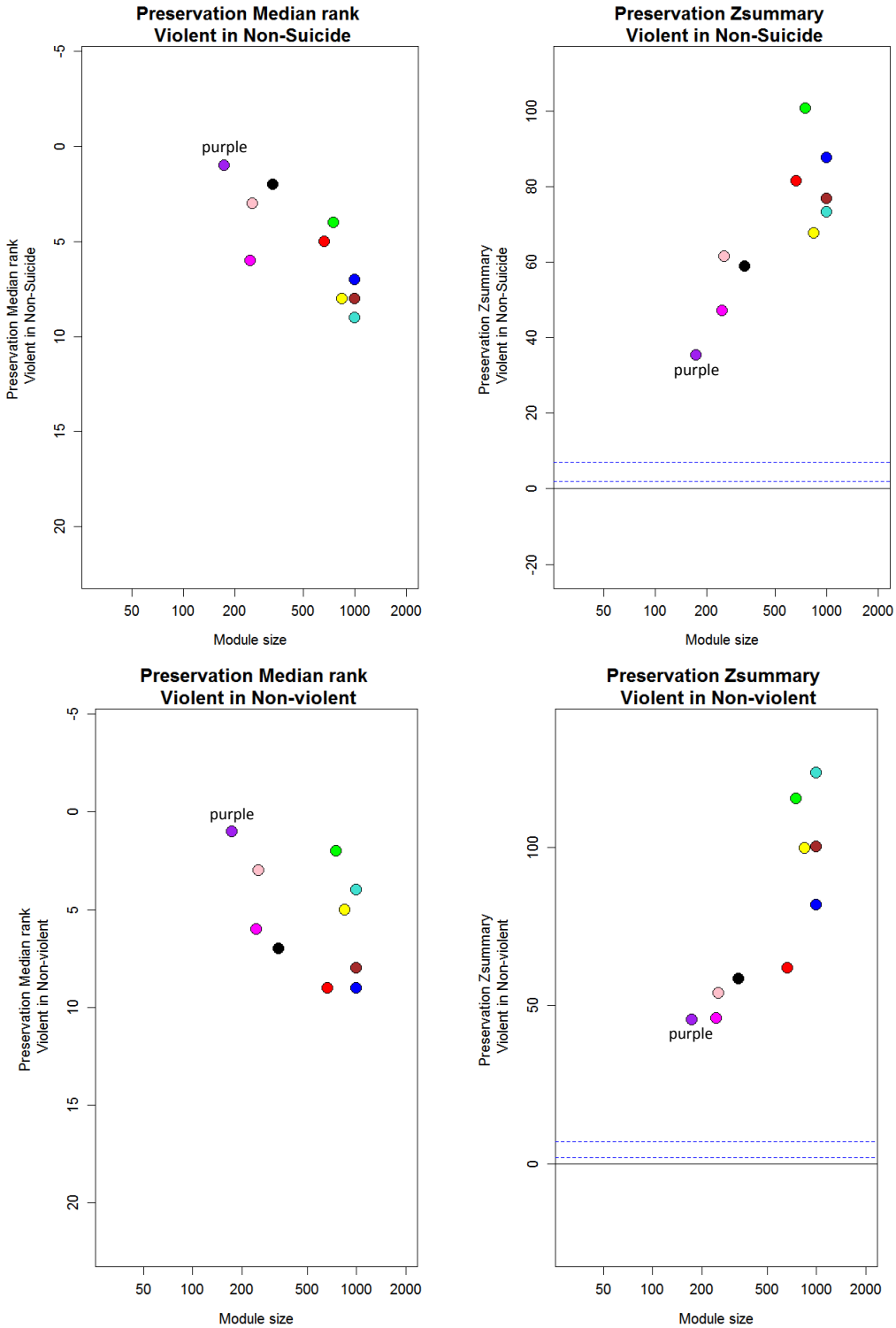
SF3: Boxplots of the effect of suicide by violent means *vs* non-suicide on the DLPFC expression of *LINC01268* in the total sample (N=228, all PZ), divided by sex. Statistics for LM (with age and RIN as covariates) in F: t=4.407 p=4.9e-05; in M: t=5.104 p=1.33e-06. *LINC01268* expression is greater in suicide by violent means compared with non-suicidal deceased in both sexes. F= female, M= male.

Supplementary Figure 4



SF4: Boxplots of the effect of suicide by violent means *vs* non-suicide on the DLPFC expression of *LINC01268* in the total sample (N=228, all PZ), divided by diagnosis. Statistics for LM (with age, sex and RIN as covariates) in SCZ: t=5.333, p=1.44e-06; in MDD: t=4.715, p=1.04e-05; in BPD: t=1.343, p=0.192. *LINC01268* expression is greater in suicide by violent means compared with non-suicidal deceased in each diagnosis; the effect is not significant in BPD, likely due to the smaller sample size.

Supplementary Figure 5



SF5: Preservation of suicide by violent means modules in (top) non-suicide and (bottom) suicide by non-violent means. Dotted lines on the Zsummary plots (right) represent the cut-off for preservation significance (i.e. first line =2 and second =10), while the median rank (left) is the measure of preservation accounting for the module size (number of gene). The 'purple' module from the suicide by violent means network is the most preserved module in the other two groups: $Z_{summary}$>30 and median rank≈1.

Captions for additional file ST2

ST2: GO terms enriched purple module. Reported are the biological processes (BP) terms showing statistically significant enrichment ($p<0.01$ after Benjamini-Hochberg correction) for the purple module. The purple module, containing *LINC01268*, was significantly enriched for 497 GO BP identities related to immunological functions such as positive regulation of immune response.

Supplementary Text

Additional Sensitivity Analyses

To minimize artifacts such as occult RNA quality differences that may affect the results and partially explain the differences between the three groups, we repeated the analysis with the addition of adjusting covariates from a PC analysis. After selecting the genes with mean expression >0.01 RPKM, a principal component (PC) analysis was performed on the whole DLPFC RNA-sequencing transcriptome for the 228 patients, using the function "*prcomp*" in 'R'. The first five components were the most relevant in terms of variance explained (square roots of the eigenvalues of the covariance/correlation matrix: 23.858, 17.500, 14.505, 11.479 and 9.007 respectively) and were employed as adjusting variables in the comparisons between groups in such analysis. Correcting for such PCs did not compromise the results: non-suicide vs suicide t=3.705, p=0.000267; non-suicide vs suicide by violent means t=5.702, p=5.09e-08; suicide by non-violent vs suicide by violent means t=-3.593, p=0.000475; non-suicide vs suicide by non-violent means t=0.268, p=0.78925. These PCs correlated with all of our covariates of interest (RIN, sex, age as well as diagnosis) and, notably, to the degree of physical trauma (e.g., PC1: t=5.667, p=4.48e-08; by factor violent vs non-violent death, regardless of suicidal intent). As a consequence, the differences in *LINC01268* levels do not seem related to injuries resulting from the violent action. We addressed this further in an analysis of a sample of non-suicide subjects, comparing *LINC01268* expression in natural deaths (mostly cardiac) and violent, non-suicidal, deaths (homicides and accidents – mostly motor vehicle accidents). These groups did not differ in *LINC01268* levels (t=1.050, p=0.3, following PC correction). These data, together with previous evidence that *LINC01268* is not influenced by the acute exposure to chemical substances(17), such as those employed for self-poisoning, support the conclusion that the lincRNA is related to the employment of violent means for suicide *per se*.

Supplementary Methods

Gene Expression

RNA extraction and sequencing. Post-mortem tissue homogenates of dorsolateral prefrontal cortex gray matter (DLPFC) approximating BA46/9 were obtained from all brains. Total RNA was extracted from ~100 mg of tissue using the RNeasy kit (Qiagen) according to the manufacturer's protocol. The poly(A)-containing RNA molecules were purified from 1 μg DNase-treated total RNA and, following purification, fragmented into small pieces using divalent cations under elevated

temperature. Reverse transcriptase and random primers were used to copy the cleaved RNA fragments into first-strand cDNA, and the second-strand cDNA was synthesized using DNA polymerase I and RNase H. We performed the sequencing library construction using the TruSeq RNA Sample Preparation v2 kit by Illumina. An index/barcode was inserted into Illumina adapters, allowing samples to be multiplexed in one lane of a flow cell. These products were then purified and enriched with PCR to create the final cDNA library for high throughput DNA sequencing using an Illumina HiSeq 2000.

RNA sequencing data processing. The Illumina Real Time Analysis (RTA) was applied to perform image analysis and base calling; the BCL Converter (CASAVA v1.8.2) followed to generate FASTQ files. The sequencing depth was over 80 million (40 million paired-end) mappable sequencing reads. Pair-end reads of cDNA sequences were aligned to the human genome (UCSC hg19) by the spliced read mapper TopHat (v2.0.4), using the reference transcriptome (Ensembl Build GRCh37.67) to initially guide the alignment. Gene counts were generated using the featureCounts tool (v1.4.3-p1) based on the more recent Ensembl v75, which was the last stable release for the hg19 genome build. Counts were converted to RPKM values using the total number of aligned reads across the autosomal and sex chromosomes (dropping reads mapping to the mitochondria chromosome). More details of this method have been described elsewhere(1).

fMRI

fMRI data acquisition. Blood Oxygen Level Dependent (BOLD) fMRI was performed on a GE Signa 3T scanner (gradient echo-planar imaging sequence, TR/TE = 2000/28 ms; 24 interleaved slices, thickness = 4 mm, gap = 1 mm; voxel size 3.75 × 3.75 × 5 mm; scan repetitions = 180; flip angle = 90°; field of view = 24 cm; matrix = 64 × 64) while subjects performed the task. The first four scans were discarded to allow for signal saturation. Stimuli were presented via a back-projection system and responses were recorded through a fiber optic response box, which allowed measurement of behavioral data as percent of correct responses and reaction time (RT). fMRI responses were modeled using a canonical hemodynamic response function and temporally filtered using a high-pass filter of 128 Hz to minimize scanner drift.

fMRI data analysis. Analysis of the fMRI data was completed using Statistical Parametric Mapping (SPM8; www.fil.ion.ucl.ac.uk/spm). Images, for each subject, were realigned to the first volume in the time series and movement parameters were extracted to exclude subjects with excessive head motion (> 2 mm of translation, > 1.5° rotation). Images were then re-sampled to a 2 mm isotropic voxel size, spatially normalized into a standard stereotactic space (Montreal Institute on Neurology, MNI, template) and smoothed using a 4 mm full-width half-maximum isotropic Gaussian kernel to minimize noise and to account for residual inter-subject differences. A boxcar model convolved with the hemodynamic response function (HRF) at each voxel was modeled. Vectors were created for angry, happy, fearful and neutral faces. Residual movement was modeled as regressor of no interest. Predetermined condition effects at each voxel were created using a t statistic, producing a statistical image for BOLD responses to brain processing of stimuli representative of each condition, i.e., angry, happy, fearful and neutral faces versus fixation crosshairs. Individual contrasts of angry faces versus crosshairs were then entered in second-level random effects models; a one-sample t-test was performed to obtain a whole sample activation map for the angry condition. A factorial regression analysis with rs7747961 genotype and

Adult BG scores as predictor variables was performed to investigate their interaction on the activity of a PFC region of interest, defined by the map of the main effect of task obtained with the one-sample t-test. For group statistics we used a statistical threshold of $P<0.05$, minimum cluster size [k]=20, family-wise error (FWE) corrected using as volume of interest the PFC, specifically the DLPFC as identified with the Wake Forest University Pickatlas (www.fmri.wfubmc.edu/software/PickAtlas). This region was chosen a priori based on previous literature suggesting a potential role in emotional processing, aggressive behaviors and suicide as discussed in main text. Outside this ROI, no activation encompassed the threshold of FWE $P<0.05$ whole-brain corrections. Finally, BOLD responses were extracted from significant clusters using MarsBar (www.marsbar.sourceforge.net/).

WGCNA

For WGCNA(2), gene level RNA-seq data from the same samples of DLPFC tissue from patients were used. Only genes with sufficient abundance (median RPKM ≥ 0.1) were retained for analysis; this criterion yielded a number of 23,172 genes. Gene expression data were adjusted to remove the unwanted variance potentially explained by RNA quality as noted above. Specifically, expression data, normalized by log2 transformation with an offset of 1 to avoid issues with 0s, were adjusted to control for unwanted variance by using the *empiricalBayesLM* function implemented in WGCNA(3). Such method removes variation due to unwanted covariates, while preserving covariates of interest, and employs empirical Bayes-moderated linear regression in a robust version, resistant to expression outliers. “Removed” covariates included RIN, PMI and pH. Since co-expression analysis can be more sensitive to variables associated with RNA quality than single-gene models, we applied a similar strategy to a previous reference(4). More in details, we first tested if our RNA quality measures (i.e. RIN, PMI, pH) were associated with the PCs derived from the expression data and used in the additional sensitivity analysis (see above). Since all of them were correlated with the PCs (not shown), we used them in the procedure of removing the unwanted covariance. On the other hand, we were interested in preserving the effect of age and sex on the co-expression network; therefore we retained them in the model. The output of the empirical Bayes adjustment was a new expression dataset that was used in the co-expression network analysis. We would mention however that removing unwanted variance of observed and latent variables with concomitant preservation of signal of interest remains an ongoing work(5); we applied here the empirical Bayes linear regression method, which has been successfully tested in previous studies(6), acknowledging that type I errors may not be fully controlled by this method.

After creating the new, adjusted expression dataset, a weighted gene co-expression analysis was performed with functions implemented in the WGCNA package(3). A co-expression network based on the samples of suicide by violent means, irrespective of diagnosis, (N=77) was created. WGCNA uses correlation between pairs of genes to construct co-expression modules. These modules can then be summarized by the “eigengene” for each module (ME). Biological inference can be drawn from the genes in these constructed modules by using gene-set enrichment analyses and by correlating module eigengenes with biological covariates.

WGCNA *blockwiseModules* function was used for automatic network construction and module detection. Parameters of network construction included robust WGCNA (bi-weight mid-correlation) with “signed” networks to allow for potentially non-linear correlations between genes that may better reflect the underlying biology of the brain.

The “Soft Thresholding” method (implemented through *pickSoftThreshold* function) was used to improve the sensitivity of network construction; this procedure applies the power function to the gene correlations prior to network construction. Results from this function help to select the adequate power to which the similarity will be raised and thus to obtain the weighted networks. For the specific networks constructed here, a power of 14 corresponding to an R^2 of 0.8 was selected. After network construction, modules of co-expression were detected with hierarchical clustering using a measure of dissimilarity (the topological overlap). Module eigengenes were calculated for identified modules of co-expression. These can be regarded as expression profiles that best characterize the gene correlations within modules. Their major advantage is the dimensionality data reduction, which makes them particularly suitable for correlation with traits of interest by eliminating the problem of multiple comparisons corrections.

Module preservation analysis was use to confirm “violent suicide” co-expression modules in expression data from “non-violent suicide” donors and from “non-suicide” donors. This analysis is implemented by *modulePreservation* function in WGCNA, which computes pair-wise module preservation statistics between a reference set (in this case the “violent suicide” expression data) and a test set (“non-violent suicide” or “non-suicide” datasets). Measures of preservation are the Z summary score that indicates no preservation for a range between 0-2, weak preservation between 2-10 and strong preservation >10, and Z median rank that orders the preserved modules accounting also for their size (number of genes).

We finally performed a Gene Ontology (GO) analysis with *enrichGO* function implemented in clusterProfiler R package(7), looking for enrichment of the “violent suicide” module in meaningful biological processes. This function performs a hypergeometric test to assess the over-representation of GO biological processes (BP) in the gene sets that constitute the modules of the violent suicide network.

SUPPLEMENTARY REFERENCES

1. Jaffe AE, Straub RE, Shin JH, Tao R, Gao Y, Collado Torres L, Kam-Thong T, Xi HS, Quan J, Chen Q, Colantuoni C, Ulrich WS, Maher BJ, Deep-Soboslay A, Cross A, Brandon NJ, Leek JT, Hyde TM, Kleinman JE, Weinberger DR. Developmental And Genetic Regulation Of The Human Cortex Transcriptome In Schizophrenia. bioRxiv. 2017.
2. Horvath S. Weighted Network Analysis: Applications in Genomics and Systems Biology. Weighted Network Analysis: Applications in Genomics and Systems Biology. 2011:1-421.
3. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
4. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM, Ruderfer DM, Oh EC, Topol A, Shah HR, Klei LL, Kramer R, Pinto D, Gumus ZH, Cicek AE, Dang KK, Browne A, Lu C, Xie L, Readhead B, Stahl EA, Xiao J, Parvizi M, Hamamsy T, Fullard JF, Wang YC, Mahajan MC, Derry JM, Dudley JT, Hemby SE, Logsdon BA, Talbot K, Raj T, Bennett DA, De Jager PL, Zhu J, Zhang B, Sullivan PF, Chess A, Purcell SM, Shinobu LA, Mangravite LM, Toyoshiba H, Gur RE, Hahn CG, Lewis DA, Haroutunian V, Peters MA, Lipska BK, Buxbaum JD, Schadt EE, Hirai K, Roeder K, Brennand KJ, Katsanis N, Domenici E, Devlin B, Sklar P. Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nature neuroscience. 2016;19:1442-1453.
5. Parsana P, Ruberman C, Jaffe AE, Schatz MC, Battle A, Leek JT. Addressing confounding artifacts in reconstruction of gene co-expression networks. bioRxiv. 2017.
6. Jacob L, Gagnon-Bartsch JA, Speed TP. Correcting gene expression data when neither the unwanted variation nor the factor of interest are observed. Biostatistics. 2016;17:16-28.
7. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284-287.